

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Tetsuya SAKAI

GAU:

SERIAL NO: NEW APPLICATION

EXAMINER:

FILED: HERewith

FOR: SYSTEM, METHOD, AND PROGRAM PRODUCT FOR QUESTION ANSWERING

REQUEST FOR PRIORITY

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

☐ Full benefit of the filing date of U.S. Application Serial Number , filed , is claimed pursuant to the provisions of 35 U.S.C. §120.

☐ Full benefit of the filing date(s) of U.S. Provisional Application(s) is claimed pursuant to the provisions of 35 U.S.C. §119(e): Application No. Date Filed

☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:

COUNTRY
Japan

APPLICATION NUMBER
2002-284328

MONTH/DAY/YEAR
September 27, 2002

Certified copies of the corresponding Convention Application(s)

☒ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

☐ were filed in prior application Serial No. filed

☐ were submitted to the International Bureau in PCT Application Number
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.

☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and

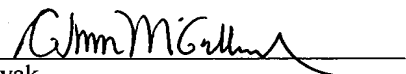
☐ (B) Application Serial No.(s)

☐ are submitted herewith

☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.


Marvin J. Spivak

Registration No. 24,913

Customer Number

22850

Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 05/03)

C. Irvin McClelland
Registration Number 21,124

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2002年 9月27日

出 願 番 号
Application Number:

特願2002-284328

[ST.10/C]:

[JP 2002-284328]

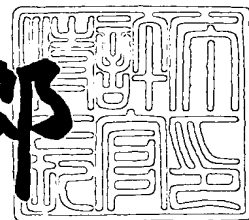
出 願 人
Applicant(s):

株式会社東芝

2003年 1月31日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3002798

【書類名】 特許願

【整理番号】 A000203682

【提出日】 平成14年 9月27日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 7/00

【発明の名称】 質問応答システム、質問応答方法、質問応答プログラム

【請求項の数】 12

【発明者】

 【住所又は居所】 神奈川県川崎市幸区小向東芝町 1 番地 株式会社東芝研
 究開発センター内

 【氏名】 酒井 哲也

【特許出願人】

 【識別番号】 000003078

 【氏名又は名称】 株式会社 東芝

【代理人】

 【識別番号】 100058479

 【弁理士】

 【氏名又は名称】 鈴江 武彦

 【電話番号】 03-3502-3181

【選任した代理人】

 【識別番号】 100084618

 【弁理士】

 【氏名又は名称】 村松 貞男

【選任した代理人】

 【識別番号】 100068814

 【弁理士】

 【氏名又は名称】 坪井 淳

【選任した代理人】

 【識別番号】 100092196

【弁理士】

【氏名又は名称】 橋本 良郎

【選任した代理人】

【識別番号】 100091351

【弁理士】

【氏名又は名称】 河野 哲

【選任した代理人】

【識別番号】 100088683

【弁理士】

【氏名又は名称】 中村 誠

【選任した代理人】

【識別番号】 100070437

【弁理士】

【氏名又は名称】 河井 将次

【手数料の表示】

【予納台帳番号】 011567

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 質問応答システム、質問応答方法、質問応答プログラム

【特許請求の範囲】

【請求項 1】 ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源を有する第 2 の知識データベースを用いて回答を求める質問応答システムにおいて、

前記質問について、前記第 1 の知識データベースを検索して第 1 の言語の回答候補を得る手段と、

前記質問を第 2 の言語に機械翻訳する手段と、

前記第 2 の言語に翻訳された質問について、前記第 2 の知識データベースを検索して第 2 の言語の回答候補を得る手段と、

前記第 2 の言語の回答候補を第 1 の言語に機械翻訳する手段と、

前記第 1 の言語の回答候補および前記第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けする手段と、
を具備することを特徴とする質問応答システム。

【請求項 2】 前記順位付けに基づいて前記回答候補の中からいずれか一つの回答を決定する手段を具備することを特徴とする請求項 1 に記載の質問応答システム。

【請求項 3】 前記第 1 および第 2 の知識データベースにおける検索ヒット件数の多寡を前記基準とすることを特徴とする請求項 1 に記載の質問応答システム。

【請求項 4】 前記回答候補の各々の簡潔さ又は網羅度を字句処理により決定する手段を具備し、該簡潔さ又は網羅度を前記基準とすることを特徴とする請求項 1 に記載の質問応答システム。

【請求項 5】 ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源を有する第 2 の知識データベースを用いて回答を求める質問応答方法において、

前記質問について、前記第 1 の知識データベースを検索して第 1 の言語の回答候補を得るステップと、

前記質問を第 2 の言語に機械翻訳するステップと、

前記第 2 の言語に翻訳された質問について、前記第 2 の知識データベースを検索して第 2 の言語の回答候補を得るステップと、

前記第 2 の言語の回答候補を第 1 の言語に機械翻訳するステップと、

前記第 1 の言語の回答候補および前記第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けするステップと、
を具備することを特徴とする質問応答方法。

【請求項 6】 前記順位付けに基づいて前記回答候補の中からいずれか一つの回答を決定するステップを具備することを特徴とする請求項 5 に記載の質問応答方法。

【請求項 7】 前記第 1 および第 2 の知識データベースにおける検索ヒット件数の多寡を前記基準とすることを特徴とする請求項 5 に記載の質問応答方法。

【請求項 8】 前記回答候補の各々の簡潔さ又は網羅度を字句処理により決定するステップを具備し、該簡潔さ又は網羅度を前記基準とすることを特徴とする請求項 5 に記載の質問応答方法。

【請求項 9】 ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源を有する第 2 の知識データベースを用いて回答を求める質問応答プログラムにおいて、

前記質問について、前記第 1 の知識データベースを検索して第 1 の言語の回答候補を得る手順と、

前記質問を第 2 の言語に機械翻訳する手順と、

前記第 2 の言語に翻訳された質問について、前記第 2 の知識データベースを検索して第 2 の言語の回答候補を得る手順と、

前記第 2 の言語の回答候補を第 1 の言語に機械翻訳する手順と、

前記第 1 の言語の回答候補および前記第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けする手順と、
をコンピュータに実行させる質問応答プログラム。

【請求項 10】 前記順位付けに基づいて前記回答候補の中からいずれか一つの回答を決定する手順を具備することを特徴とする請求項 9 に記載の質問応答

プログラム。

【請求項 1 1】 前記第 1 および第 2 の知識データベースにおける検索ヒット件数の多寡を前記基準とすることを特徴とする請求項 9 に記載の質問応答プログラム。

【請求項 1 2】 前記回答候補の各々の簡潔さ又は網羅度を字句処理により決定する手順を具備し、該簡潔さ又は網羅度を前記基準とすることを特徴とする請求項 9 に記載の質問応答プログラム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、ユーザの入力した質問に対する回答を出力する質問応答システム（question answering system）に関する。

【0 0 0 2】

【従来の技術】

インターネットの検索エンジンに代表されるように、ユーザの検索要求に適合する文書を検索してランキングする文書検索（document retrieval）技術は広く普及している。しかし、文書検索は、「．．．に関する新聞記事が読みたい」、「．．．に関する Web ページが見たい」といった検索要求を満足することはできるが、「〇×社の社長は誰か？」、「富士山の高さは？」、「鯨は絶滅しかかっているか？」といった質問に対する答えを回答することができない。すなわち、文書検索は文書あるいは文書中のパッセージを回答するものであるに過ぎず、ユーザは文書検索の出力結果から自分で答えを探しださなくてはならない。

【0 0 0 3】

入力された質問に対する答えを出力するものとして、質問応答システム（question answering system）がある。質問応答システムは、例えば、「〇×社の社長は誰か？」という質問に対し、〇×社のホームページなど〇×社に関する文書を出力するのではなく、〇×社の社長の人名といった答えを出力する。また、「富士山の高さは？」のような質問に対しては、「富士山は 3 7 7 6 m です。」のような答えを出力する。質問応答システムは、例えば非特許文献 1 に見られるよ

うに、古くからエキスパートシステム的一种として研究が行われてきたが、近年、情報検索 (information retrieval) や情報抽出 (information extraction) などの研究の発展形として新たに注目を集めている。

【 0 0 0 4 】

例えば日本語の質問を受け付けて、日本語の知識源を用いて回答を生成する単言語質問応答システム (monolingual question answering system) は、既存の情報検索技術 (特定の語を含むテキストを探し出す技術) および情報抽出技術 (例えば人名、地名、数値などの特定の種類の情報を抜き出す技術) の組み合わせ利用により、ある程度は容易に実現できる。しかし、単言語の知識源を用いて回答を生成する単言語質問応答システムには以下のような問題点がある。

【 0 0 0 5 】

第 1 の問題点は、質問に対する回答を作成するために必要な情報が量的に十分でないことである。これは、回答のカバレッジ (coverage) および信頼性 (reliability) の低下につながる。例えば、ある日本語の質問に対して回答するのに必要な情報が英語の web (ウェブ) ページには記載されているが日本語の web ページには記載されていない場合であり、このような英語情報を活用できない日本語単言語質問応答システム (Japanese monolingual question answering system) は回答作成に失敗する。これはカバレッジの問題である。また、例えば、「〇×社の社長は誰か？」という質問に対し、日本語知識源からは「〇×社の社長は A 氏である」、「〇×社の社長は B 氏である」という 2 つの回答候補が検索可能であって、また英語知識源からは "The president of 〇× Corporation is Mr. A." という 1 つの回答候補が検索可能であるような場合、日本語知識源しか活用できない日本語単言語質問応答システムは、A 氏と B 氏のどちらが回答として信頼性が高いかを判定することができない。実際には、日本語知識と英語知識を総合すると A 氏の方が回答としての信頼性が高いと考えられる。なお、質問応答システムとは異なる技術であるが、検索対象データベースの記述言語と入力キーワードの記述言語とが異なる場合においても、入力キーワードに忠実な検索結果の出力を得ることができる情報検索装置が知られている (例えば特許文献 1 参照。) 。

【0006】

第2の問題点は、質問に対する回答を作成するために必要な情報が質的に偏ってしまうことである。例えば、「鯨は絶滅しかかっているか？」との質問に対して、捕鯨が行われている国の言語で書かれたwebページのみを知識源として用いた場合、「鯨は絶滅しかかっているはいない。むしろある種の鯨は増殖している。」といった内容のみの回答しか得られず、逆に、捕鯨を禁止あるいは反対している国の言語で書かれたwebページのみを知識源として用いた場合、「鯨は捕鯨国の乱獲のせいで絶滅しかかっている。」といった内容のみの回答のみしか得られないかも知れない。このように、言語を限定することは、本来的に多様であるべき観点をも限定することにつながる可能性がある。

【0007】

第3の問題点は、言語毎に知識源の充実度が異なる点である。知識源の充実度が異なることは、ある特定の質問に対しては、この質問への回答が充実している言語Aの知識源を用いることが好ましく、また、他の特定の質問に対しては言語Aではなくこの質問に対する回答が充実している言語Bの知識源を用いることが好ましいというケースは多々起こり得る。例えば、エリザベス女王に関する質問に対しては英語のwebページが最も充実した知識源かも知れないが、相撲に関する質問に関しては日本語のwebページが最も充実した知識源かも知れないというケースである。このような充実度の違いに対処できない単言語質問応答システムでは、質問によって回答の質が大きくばらついてしまう。

【0008】

【非特許文献1】

Wendy G. Lehnert: "The Process of Question Answering - A Computer Simulation of Cognition", Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 1978

【0009】

【特許文献1】

特開平11-219368号公報

【0010】

【発明が解決しようとする課題】

本発明はかかる事情を考慮してなされたものであり、ユーザが入力した質問に対する回答を出力する質問応答システムにおいて、言語が異なる複数の知識源を活用し、これにより回答のカバレッジ、信頼性、多様性、安定性を向上することを目的とする。

【0 0 1 1】

【課題を解決するための手段】

本発明に係る質問応答システムは、ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源を有する第 2 の知識データベースを用いて回答を求める質問応答システムにおいて、前記質問について、前記第 1 の知識データベースを検索して第 1 の言語の回答候補を得る手段と、前記質問を第 2 の言語に機械翻訳する手段と、前記第 2 の言語に翻訳された質問について、前記第 2 の知識データベースを検索して第 2 の言語の回答候補を得る手段と、前記第 2 の言語の回答候補を第 1 の言語に機械翻訳する手段と、前記第 1 の言語の回答候補および前記第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けする手段と、を具備することを特徴とする質問応答システムである。

【0 0 1 2】

上記質問応答システムの構成において、さらに、前記順位付けに基づいて前記回答候補の中からいずれか一つの回答を決定する手段を具備してもよい。

【0 0 1 3】

また、前記第 1 および第 2 の知識データベースにおける検索ヒット件数の多寡を前記基準としてもよい。

【0 0 1 4】

また、前記回答候補の各々の簡潔さ又は網羅度を字句処理により決定する手段を具備し、該簡潔さ又は網羅度を前記基準としてもよい。

【0 0 1 5】

本発明に係る質問応答方法は、ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源

を有する第 2 の知識データベースを用いて回答を求める質問応答方法において、前記質問について、前記第 1 の知識データベースを検索して第 1 の言語の回答候補を得るステップと、前記質問を第 2 の言語に機械翻訳するステップと、前記第 2 の言語に翻訳された質問について、前記第 2 の知識データベースを検索して第 2 の言語の回答候補を得るステップと、前記第 2 の言語の回答候補を第 1 の言語に機械翻訳するステップと、前記第 1 の言語の回答候補および前記第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けするステップと、を具備することを特徴とする質問応答方法である。

【 0 0 1 6 】

上記質問応答方法において、前記順位付けに基づいて前記回答候補の中からいずれか一つの回答を決定するステップをさらに具備してもよい。

【 0 0 1 7 】

また、前記第 1 および第 2 の知識データベースにおける検索ヒット件数の多寡を前記基準としてもよい。

【 0 0 1 8 】

また、前記回答候補の各々の簡潔さ又は網羅度を字句処理により決定するステップをさらに具備し、該簡潔さ又は網羅度を前記基準としてもよい。

【 0 0 1 9 】

本発明に係る質問応答プログラムは、ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源を有する第 2 の知識データベースを用いて回答を求める質問応答プログラムにおいて、前記質問について、前記第 1 の知識データベースを検索して第 1 の言語の回答候補を得る手順と、前記質問を第 2 の言語に機械翻訳する手順と、前記第 2 の言語に翻訳された質問について、前記第 2 の知識データベースを検索して第 2 の言語の回答候補を得る手順と、前記第 2 の言語の回答候補を第 1 の言語に機械翻訳する手順と、前記第 1 の言語の回答候補および前記第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けする手順と、をコンピュータに実行させる質問応答プログラムである。

【 0 0 2 0 】

上記質問応答プログラムにおいて、前記順位付けに基づいて前記回答候補の中からいずれか一つの回答を決定する手順を具備してもよい。

【 0 0 2 1 】

また、前記第 1 および第 2 の知識データベースにおける検索ヒット件数の多寡を前記基準としてもよい。

【 0 0 2 2 】

また、前記回答候補の各々の簡潔さ又は網羅度を字句処理により決定する手順をさらに具備し、該簡潔さ又は網羅度を前記基準としてもよい。

【 0 0 2 3 】

【発明の実施の形態】

以下、図面を参照しながら本発明の実施形態を説明する。

【 0 0 2 4 】

図 1 は本発明の一実施形態に係る質問応答システムの概略構成を示すブロック図である。この質問応答システムは、例えば汎用の計算機（コンピュータ）および同計算機上で動作するソフトウェアを用いて実現することができ、入力部 6 および出力部 8 からなるユーザインターフェース 4、検索部 10、情報抽出部 15、回答作成部 18、翻訳部 19 から構成される。ユーザインターフェース 4 には、キーボードやマウスなどの入力装置、ディスプレイなどの出力装置等のハードウェアが用いられる。検索部 10、情報抽出部 15、回答作成部 18、翻訳部 19 は、汎用のオペレーティングシステムの下で動作するコンピュータプログラムのモジュールとして実現できる。

【 0 0 2 5 】

なお、本発明は任意数の複数言語の知識源を扱うものを含むが、実施形態の説明では、便宜上、言語 1 および言語 2 からなる 2 つの言語の知識源を扱うこととし、例えば、言語 1 は「日本語」、言語 2 は「英語」とする。

【 0 0 2 6 】

以下、先ずは本システムの全体的な処理手順を説明し、その後に主要なモジュールによる具体的な処理手順について詳述する。

【 0 0 2 7 】

（全体的な処理手順）

図 1 において、点線の矢印は質問に関する情報の流れを、実線の矢印は回答に関する情報の流れを表している。

【 0 0 2 8 】

情報抽出部 1 5 は、あらかじめ、複数言語で記述された文書 1 6， 1 7 から情報抽出を行い、言語毎に知識データベース 1 3， 1 4 を作成する。

【 0 0 2 9 】

入力部 6 に対しユーザ 2 から言語 1（ここでは日本語）の質問が入力されると、該入力された質問は検索部 1 0 および翻訳部 1 9 に渡される。翻訳部 1 9 は、質問を言語 2（ここでは英語）の質問に翻訳して検索部 1 0 に渡す。

【 0 0 3 0 】

検索部 1 0 は、入力部 6 から渡された質問について、言語 1（日本語）の知識データベース（以下、「日本語知識データベース」） 1 3 を検索し、また、翻訳部 1 9 により英語に翻訳された質問について、言語 2（英語）の知識データベース（以下、「英語知識データベース」） 1 4 を検索する。これにより得られる日本語知識データベース 1 3 の検索結果（言語 1 の回答候補）は回答作成部 1 8 に渡され、英語知識データベース 1 4 の検索結果（言語 2 の回答候補）は翻訳部 1 9 に渡される。次に、翻訳部 1 9 は、言語 2 の回答候補を言語 1 に翻訳して回答作成部 1 8 に渡す。すなわち、英語で記述された回答候補が日本語に翻訳されて回答作成部 1 8 に渡される。

【 0 0 3 1 】

以上により回答作成部 1 8 では言語 1（日本語）に統一された回答候補が得られる。さらに回答作成部 1 8 は回答候補間の比較を行い回答の順位を判定した上で回答情報を出力部 8 に渡す。以上の処理において、従来の質問応答システムと異なる重要な点は、検索結果として得られ、言語の異なる回答候補のうち少なくとも一方の言語による回答候補が翻訳部 1 9 により機械翻訳され、回答候補が他方の言語に統一されるとともに、該言語統一がなされた回答候補群に対して回答作成部 1 8 が比較処理を行う点である。

【 0 0 3 2 】

以下、これについて情報抽出部 15、検索部 10、翻訳部 19、回答作成部 18 の各々の処理手順に沿って詳細に説明する。

【0033】

(情報抽出部の処理手順)

図 2 は情報抽出部 15 の処理手順の一例を示すフローチャートである。

【0034】

情報抽出部 15 は、言語 i ($i = 1, 2, \dots$) で書かれた j 番目の文書 ($j = 1, 2, \dots$) を読み込み、該文書から既存の情報抽出技術を利用して情報抽出を行い、その結果を言語 i の知識データベースに登録する。

【0035】

ここで、情報抽出の具体的方法としては、例えば形態素解析 (morphological analysis) 及びパターンマッチングによる方法が挙げられる。例えば、知識源が日本語の場合、文書 16 中に「〇×社 (社長: 〇×太郎)」という表現を含む場合、これを形態素解析して、

「/〇×社<固有名詞>/ (<記号>/社長<一般名詞>/ : <記号>/〇×太郎<固有名詞>/) <記号>」

という解析結果が得られる。なお、「/」は品詞の区切りを表す。

【0036】

ここで、仮に、

「/X<固有名詞>/ (<記号>/社長<一般名詞>/ : <記号>/Y<固有名詞>/) <記号>」

という形態素の並びを「X [PRESIDENT == Y]」という知識表現に書き換える情報抽出ルールを用いることにより、

「〇×社 [PRESIDENT == 〇×太郎]」

という知識を得ることができる。

【0037】

また、例えば、

「/X<固有名詞>/の<助詞>/Y<固有名詞>/社長<一般名詞>」

という形態素の並びを「X [PRESIDENT == Y]」という知識表現に書

き換える情報抽出ルールを用いることにより、

「〇×社の〇×太郎社長．．．」という表現から、同様に、「〇×社 [P R E S I D E N T = = 〇×太郎] 」という知識を得ることができる。

【 0 0 3 8 】

さらに、例えば知識源が英語の場合、形態素解析に代えて品詞同定 (Part-of-Speech tagging) を行うことにより、文書 1 7 中の "Taro 〇×, president of 〇× Corporation, ..." のような表現から、例えば、

「〇×_C o r p o r a t i o n [P R E S I D E N T = = T a r o _ 〇×] 」
という表現形式の知識を得ることができる。

【 0 0 3 9 】

なお、上記のような表現形式の知識には、元となる文書の識別番号が付加されてもよい。こうすれば、各知識データがどのような文書テキストから得られたものかを後の段階で把握することが可能となる。

【 0 0 4 0 】

情報抽出部 1 5 は、以上のようにして得られた知識を、言語毎に、知識データベース 1 3, 1 4 に登録する。

【 0 0 4 1 】

(検索部の処理手順)

図 3 は検索部 1 0 の処理手順の一例を示すフローチャートである。

【 0 0 4 2 】

検索部 1 0 はまず、入力部 6 からユーザからの質問を受け取り (ステップ S 1 1)、さらに、該質問の翻訳結果を翻訳部 1 9 から受け取る (ステップ S 1 2)。そして、言語 i ($i = 1, 2, \dots$) で書かれた各質問に対して、検索条件の生成を行う。例えば、「〇×社の社長は？」という日本語の質問を、検索部 1 0 は「〇×社 [P R E S I D E N T = = *] 」という表現形式の検索条件に変換する (ステップ S 1 3)。ここで、文字「*」はワイルドカードを表す。検索部 1 0 は、生成された検索条件を用いて日本語知識データベース 1 3 を検索する (ステップ S 1 5)。これにより、例えば「〇×社 [P R E S I D E N T = = 〇×太郎] 」のようなデータがマッチし、回答候補として「〇×太郎」を得ることが

できる。なお、一般には回答候補は複数得られる。

【0043】

検索部10は、日本語以外の質問に対しても同様な処理を行う。すなわち、例えば”Who is the president of OX Corporation?”という英語の質問に対しては、これを「OX_Corporation [PRESIDENT==*]・」という検索条件に変換し（ステップS14）、これを用いて英語知識データベース14を検索する（ステップS15）。これにより、回答候補として「Tarō_OX」を得る。

【0044】

検索部10は、ステップS16において、現在処理中の質問の言語が、ユーザが入力した質問の言語と同じであるか否かを判定し、その判定結果に応じて回答候補を回答作成部18に直接渡すか（ステップS17）、翻訳部19に渡す（ステップS18）。例えば、ユーザによる質問の入力言語が日本語であったならば、日本語知識データベース13の検索により得られた回答候補はそのまま回答作成部18に渡され、英語知識データベース14の検索により得られた回答候補は日本語への翻訳のために翻訳部19に渡されることになる。

【0045】

（翻訳部の処理手順）

図4（a）は翻訳部19による質問の処理手順の一例を示すフローチャート、図4（b）は翻訳部19による回答候補の処理手順の一例をフローチャートである。翻訳部19は、質問を機械翻訳して検索部10に渡す。また、回答候補を機械翻訳して回答作成部18に渡す。

【0046】

例えば、「OX社の社長は？」との質問を入力部6から受け取ると（ステップS21）、翻訳部19はこれを”Who is the president of OX Corporation?”のように機械翻訳し（ステップS22）、該機械翻訳の結果を検索部10に渡す（ステップS23）。一方、例えば「Tarō_OX」のような回答候補の文字列を検索部10から受け取る（ステップS24）と、翻訳部19はこれを「OX 太郎」のように機械翻訳（ステップS25）し、該機械翻訳の結果を回答作成部

1 8 に渡す（ステップ S 2 6）。

【 0 0 4 7 】

（回答作成部の処理手順）

図 5 は本実施形態の回答作成部 1 8 の処理手順の一例を示すフローチャートである。

【 0 0 4 8 】

回答作成部 1 8 は、まず検索部 1 0 から回答候補を受け取り（ステップ S 2 7）、次に翻訳部 1 9 から回答候補を受け取る（ステップ S 2 8）。上述したように、検索部 1 0 から受け取る回答候補の言語と、翻訳部 1 9 から受け取る回答候補の言語は同一である。例えば、ユーザが日本語で質問をした場合、検索部 1 0 から受け取る回答候補は日本語知識データベース 1 3 の検索により得られた日本語の回答候補そのものであり、一方、翻訳部 1 9 から受け取る回答候補は、検索部 1 0 が英語知識データベース 1 4 を検索して得られた英語の回答候補を日本語に翻訳したものである。このように、回答作成部 1 8 は単一言語のみを扱う。

【 0 0 4 9 】

回答作成部 1 8 はこれら回答候補同士の比較処理を行う（ステップ S 2 9）。これにより回答の順位を決定し、最適な回答あるいはランク付けがなされた回答を出力部 8 に渡す（ステップ S 3 0）。以下、回答の順位判定方法について詳述する。

【 0 0 5 0 】

（回答の順位決定方法）

再度、「〇×社の社長は？」という日本語の質問が入力された場合を考える。ここで、「情報抽出部の処理手順」で述べたように「／X<固有名詞>／の<助詞>／Y<固有名詞>／社長<一般名詞>」という形態素の並びを「X [P R E S I D E N T = = Y] 」という知識表現に書き換える情報抽出ルールを利用しているものとし、日本語知識データベース 1 3 の作成に用いた日本語文書 1 6 中に

- （a） 「〇×社の〇×太郎社長」
- （b） 「〇×社の〇×社長」

(c) 「〇×社は．．．△△社への出資を決めた。〇×社の△△社長に対する期待は大きい。」

という表現が含まれていたと仮定する。

【0051】

回答候補としては「〇×太郎」「〇×」「△△」などが得られる。ここで、「△△」という回答候補は、上記(c)の「〇×社の△△社長（に対する期待は大きい）」という表現に、情報抽出ルールがマッチしたために得られてしまったが、実際には回答として妥当ではないものと仮定する（なお、情報抽出の精度が高くても、もとの文書自体に真実でないことが書かれている場合も考えられるので、一般に、回答候補の中には妥当でないものが混在する可能性は少なくない）。

【0052】

ここでは、日本語知識データベース13を検索した結果、「〇×太郎」という回答候補が3件、「〇×」という回答候補が1件、「△△」という回答候補が1件得られたとする。また、「〇×社の社長は？」という日本語の質問を英語に翻訳し、該質問の英語への翻訳結果に基づいて英語知識データベース14を検索し、これにより検索された回答候補を日本語に翻訳した結果、「〇×太郎」という回答候補が2件、「〇×」という回答候補が1件得られたとする。以上の場合において、回答の順位を、例えば、単純な多数決法にしたがって決定することができる。

【0053】

図6は本実施形態の質問応答システムにより得られた回答候補の出力方法の一例を示す図である。ここでは、複数の回答（候補）1～3（「〇×太郎」、「〇×」、「△△」）が、日本語知識データベース13への検索、および英語知識データベース14への検索においてヒットした順にソートされている(202)。同図において、黒丸印「●」で示されるマーク204は、ヒットした知識データを表している。このマーク204は、表203において知識源別に区分して表示されることから、知識データの言語種別をユーザが判断できる。なお、このようなマーク表示は一例に過ぎない。例えば、マーク204に代えて文書IDなどを示してもよい。また、マーク204をクリックابلにし、ユーザのクリック指示

に応じて知識源の文書中における該当箇所を表示してもよい。

【 0 0 5 4 】

図 6 の表示例においては、回答 2 「○×」と回答 3 「△△」の日本語知識データベース 1 3 におけるヒット件数がともに 1 である。従来の単言語知識源を用いた質問応答システムでは、どちらの回答を採用すれば良いか判断できない。しかし、本発明の実施形態では、回答 2 「○×」については、日本語のみならず英語の知識源からも得られていることから、日本語の知識源のみから得られた回答 3 「△△」よりもより信頼性が高いと判定できる。

【 0 0 5 5 】

また、図 6 の表示例では、回答候補の出力方法をユーザが選択できるようにするためのチェックボックス 2 0 1 が設けられており、ここでは、「多数決」が選択されている。

【 0 0 5 6 】

出力方法の他の選択肢としては、多数決とは逆に、回答候補のユニークさ（珍しさ）を基準に順位付けして表示する「ユニーク」や、回答候補の網羅性（詳細度）を基準に順位付けして表示する「網羅性」、回答回答の簡潔さを基準に順位付けして表示する「簡潔さ」などがある。また、単純にヒット件数の多寡を基準にソートするのではなく、例えば日本語知識データベース 1 3 で 2 回ヒットした回答候補よりも、日本語知識データベース 1 3 および英語知識データベース 1 4 で 1 回ずつヒット（ヒット数の合計はともに 2 である）した回答候補を優先するような順位付けを行ってもよい。

【 0 0 5 7 】

例えば、回答候補「○×」が「○×太郎」の部分文字列であることは字句処理に基づいて容易に判定可能である。そこで、より情報量の多い「○×太郎」のほうを優先的に表示するようにしてもよい。

【 0 0 5 8 】

回答候補の順位を、網羅性あるいは簡潔さの観点から決定する別の例を図 7 に示す。ここでの質問は「酵素って何？」という、用語の定義を回答として要求する日本語の質問である（3 0 0）。このような質問 3 0 0 に対処する場合、情報

抽出部 1 5 は、例えば「. . . は. . . の一種です。」のような表現を含むテキスト（例えば文や段落）を用語定義とみなし、このような表現をあらかじめ抽出しておく。また、例えば英語の知識源に対しては、“... is a kind of ...”や、“... is a type of ...”のような慣用表現を含むテキストを定義とみなし、あらかじめ抽出しておく。

【 0 0 5 9 】

図 7 の例のように、日本語知識データベース 1 3 に対する定義表現の検索により、例えば、

A 1 : 「酵素は、触媒の一種です。触媒とは、化学反応を速める. . . 」

というテキストと、

A 2 : 「酵素は触媒の一種。」

というテキストとが回答として得られているものと仮定する。さらに、「酵素って何？」との日本語の質問を機械翻訳することにより、“What is an enzyme?”という英語の質問を得て、英語知識データベース 1 4 に対する定義表現の検索により“An enzyme is a kind of catalyst.”というテキストが回答として得られているものと仮定する。

【 0 0 6 0 】

上記英語の回答を機械翻訳により日本語に翻訳すると、例えば、A 2' 「酵素は触媒の一種です。」が得られる。したがって、回答作成部 1 8 は、検索部 1 0 から上記回答 A 1 および A 2 を、翻訳部 1 9 から A 2' を受け取ることになる。

【 0 0 6 1 】

この場合、回答作成部 1 8 は、例えば A 1, A 2 および A 2' のそれぞれを形態素解析して語の「異なり」を求め、これに基づいて回答候補の整理および優先順位づけを行うことができる。

【 0 0 6 2 】

具体的には、回答 A 1 からは「酵素、触媒、一種、化学、反応、. . . 」のような語の異なりが得られ、A 2 および A 2' からは「酵素、触媒、一種」のような語の異なりが得られる。これにより、A 2 および A 2' は回答としては等価であること、および、A 1 は A 2 および A 2' よりも網羅性（詳細度）が高いこと

がわかる。これを図 7 に示すように、回答の網羅性の高い順に、ユーザに提示する。

【0 0 6 3】

逆に、ユーザが「簡潔さ」を求める場合には、図 7 の逆順に表示を行えばよい。

【0 0 6 4】

なお、以上の説明では、回答候補に対して順位を付与し、これにもとづくソート結果をユーザに提示する場合について説明したが、上記順位が最大なものを 1 件のみ表示するようにしてもよい。

【0 0 6 5】

文書検索において機械翻訳などを利用することにより、例えば日本語検索要求で英語文書の検索を実現する言語横断検索 (cross-language information retrieval) という技術が知られているが、これはあくまで文書をランキングするために検索要求と個々の文書との類似度を算出するものであり、機械翻訳を施した上で回答候補同士の比較を行い、最適な回答を選定する本発明の実施形態とは異なる。

【0 0 6 6】

また、本発明は上述した実施形態に限定されず種々変形して実施可能である。

【0 0 6 7】

【発明の効果】

以上説明したように、本発明によれば、ユーザが入力した質問に対する回答を出力する質問応答システムにおいて、言語が異なる複数の知識源を活用し、これにより回答のカバレッジ、信頼性、多様性、安定性を向上できる。

【図面の簡単な説明】

【図 1】 本発明の一実施形態に係る質問応答システムの概略構成を示すブロック図

【図 2】 実施形態における情報抽出部の処理手順の一例を示すフローチャート

【図 3】 実施形態における検索部の処理手順の一例を示すフローチャート

【図 4】 実施形態における翻訳部の処理手順の一例を示すフローチャート

【図 5】 実施形態における回答作成部の処理手順の一例を示すフローチャート

【図 6】 実施形態における質問応答システムにより得られた回答候補の出力方法の一例を示す図

【図 7】 実施形態における質問応答システムにより得られた回答候補の出力方法の他の例を示す図

【符号の説明】

2 … ユーザ

4 … ユーザインタフェース

6 … 入力部

8 … 出力部

1 0 … 検索部

1 3 … 言語 1（日本語）の知識データベース（DB）

1 4 … 言語 2（英語）の知識データベース（DB）

1 5 … 情報抽出部

1 6 … 言語 1 の文書データ

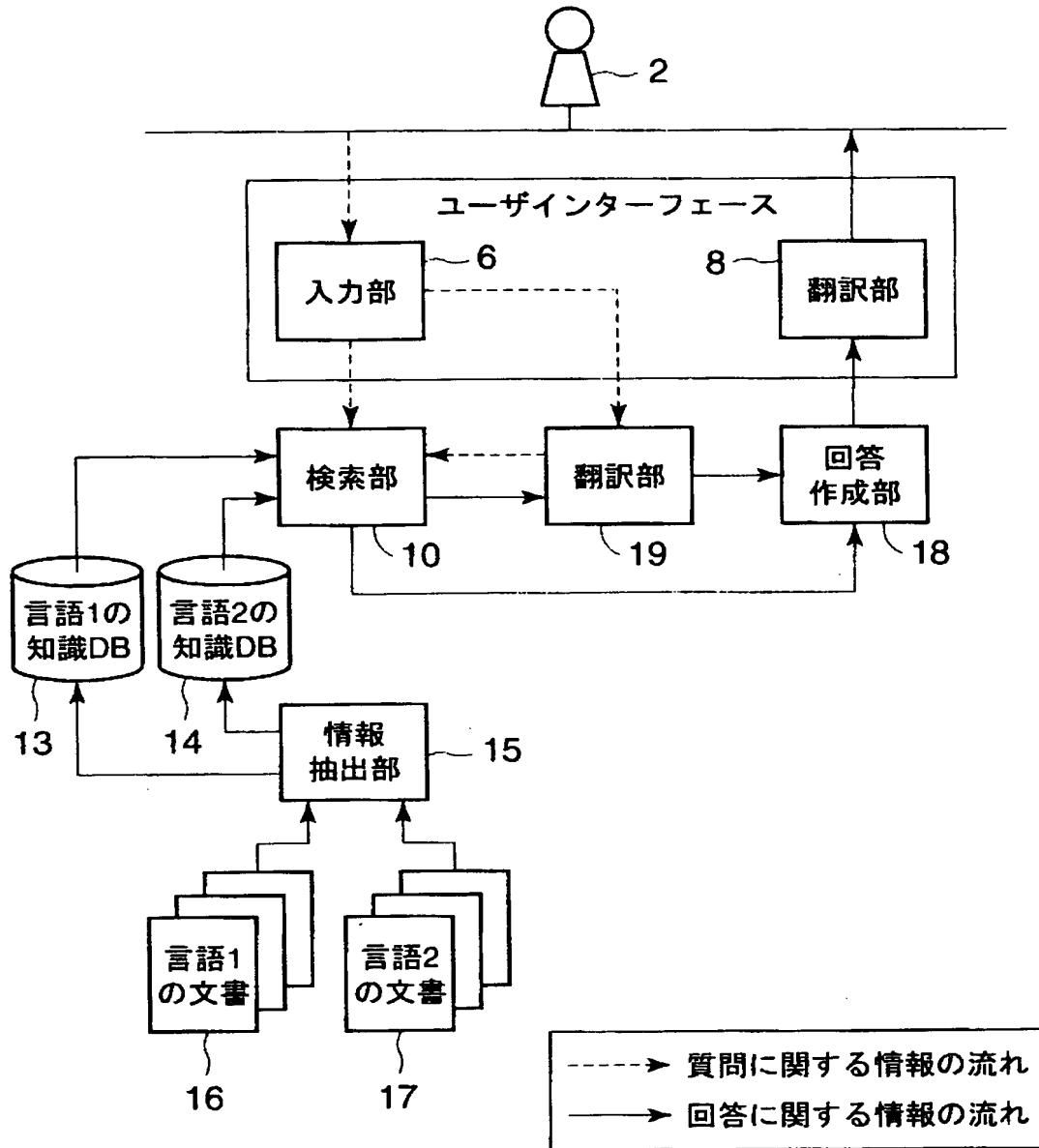
1 7 … 言語 2 の文書データ

1 8 … 回答作成部

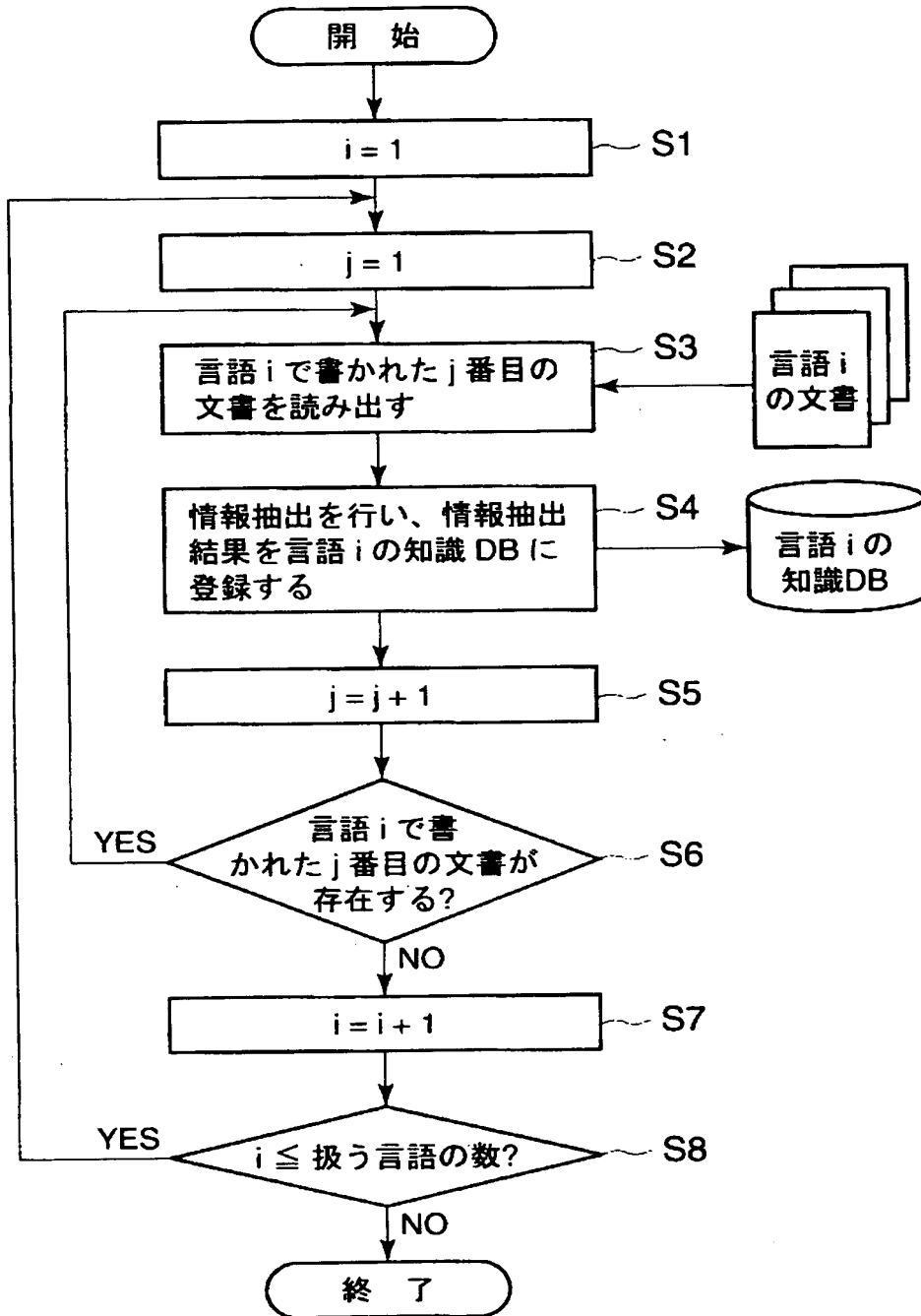
1 9 … 翻訳部

【書類名】 図面

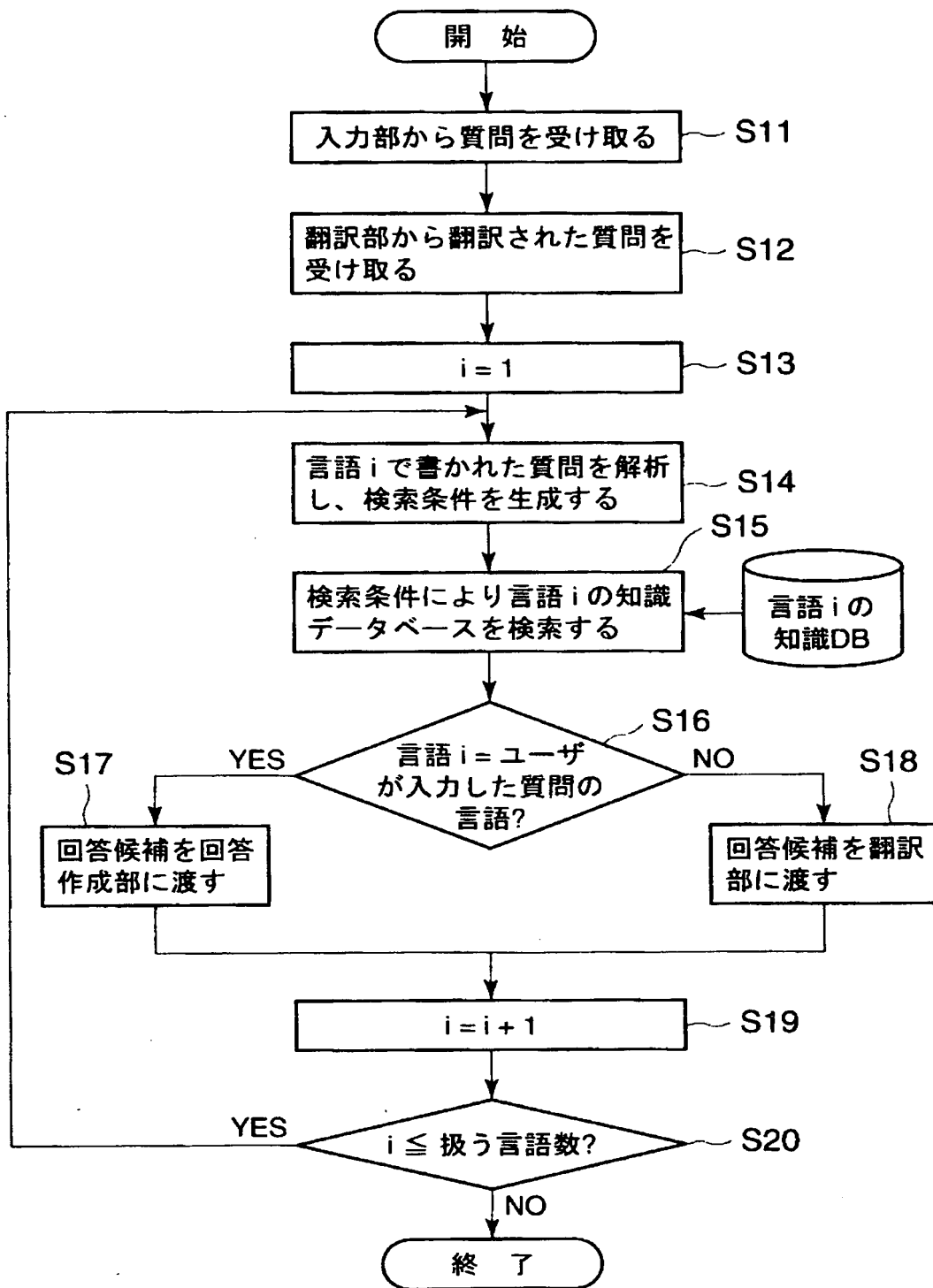
【図 1】



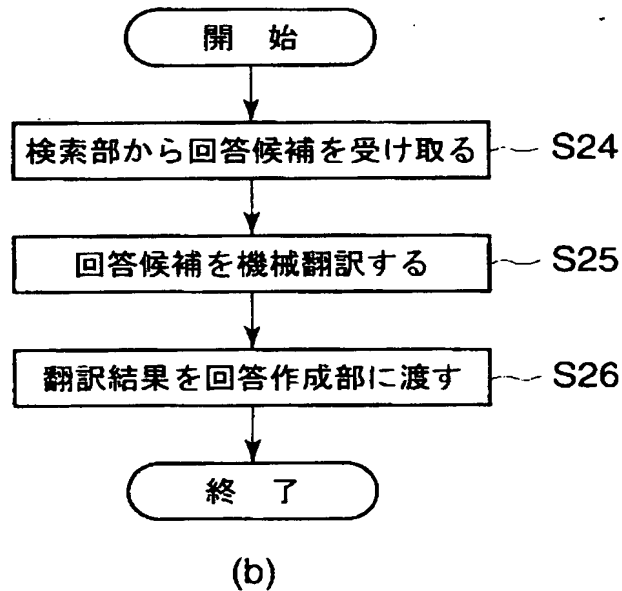
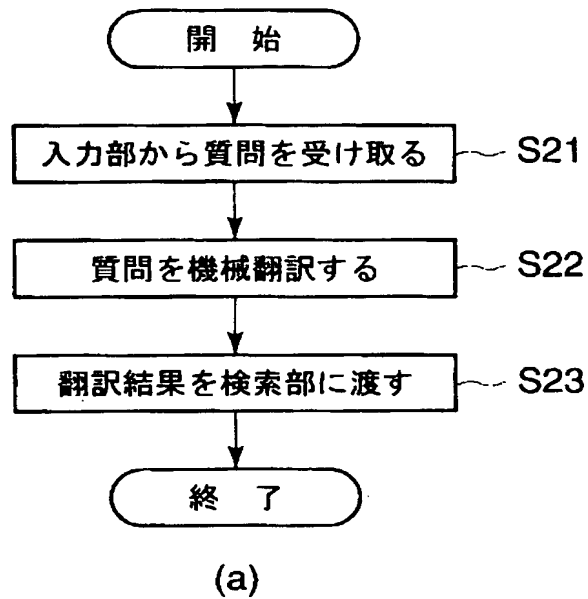
【図 2】



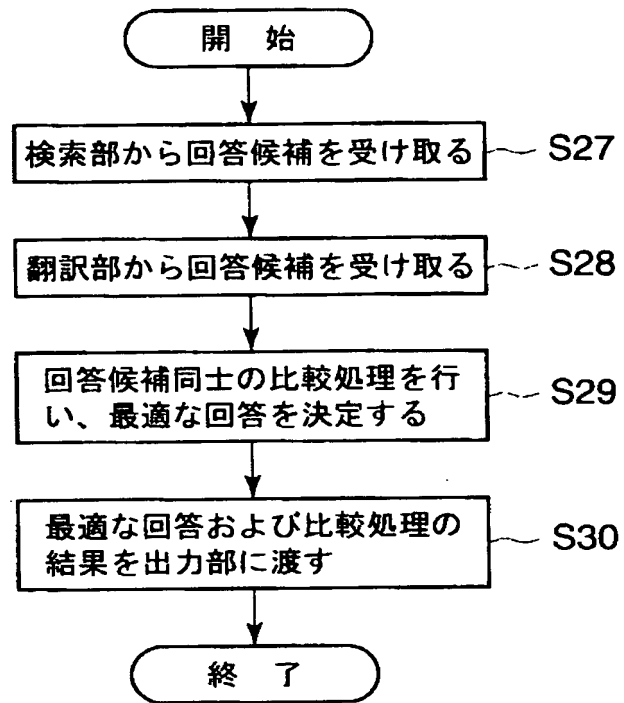
【図 3】



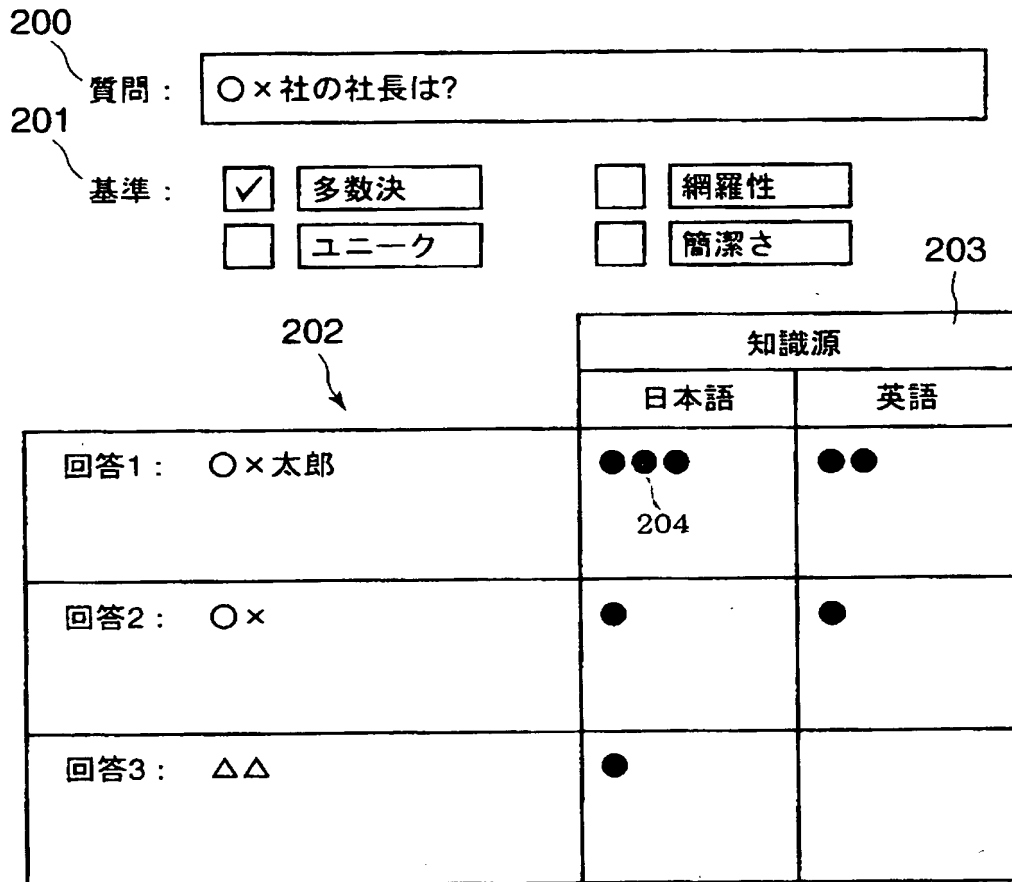
【図 4】



【図 5】



【図 6】



【図 7】

300 質問： 酸素ってなに？

301 基準： ☐ 多数決 ☒ 網羅性
☐ ユニーク ☐ 簡潔さ

302

303

		知識源	
		日本語	英語
A1	回答1： 酸素は、触媒の一種です。 触媒とは、化学反応を速める・・・	● 204	
A2	回答2： 酸素は、触媒の一種。	●	●

【書類名】 要約書

【要約】

【課題】 質問応答システムにおいて、言語が異なる複数の知識源を活用し、回答のカバレッジ、信頼性、多様性、安定性を向上すること

【解決手段】 ユーザが第 1 の言語で入力した質問について、該第 1 の言語の知識源を有する第 1 の知識データベースと、第 2 の言語の知識源を有する第 2 の知識データベースを用いて回答を求める質問応答システムにおいて、質問について、第 1 の知識データベースを検索して第 1 の言語の回答候補を得る。該質問を第 2 の言語に機械翻訳し、第 2 の知識データベースを検索して第 2 の言語の回答候補を得る。ここで、第 2 の言語の回答候補を第 1 の言語に機械翻訳する。第 1 の言語の回答候補および第 2 の言語の回答候補の第 1 の言語への機械翻訳結果の全てを、所定の基準に基づき順位付けし、ユーザに提示する。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000003078]

1. 変更年月日	2001年 7月 2日
[変更理由]	住所変更
住 所	東京都港区芝浦一丁目1番1号
氏 名	株式会社東芝